学校编码：10384　　　　分类号______密级______

学号：19020091152248　　　　UDC______

厦门大学

硕　士　学　位　论　文

参数不确定性在有马尔可夫转换的状态空间模型

Parameter uncertainty in State Space Models with Markov Switching

江　长　隆

指导教师姓名：　黄　荣　坦　副教授

专　业　名　称：　概率论与数理统计

论文提交日期：　2012 年　　月

论文答辩时间：　2012 年　　月

学位授予日期：　2012 年　　月

答辩委员会主席：__________

评　　阅　　人：__________

2012 年　　月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（　　　　　　　　　　　）课题（组）的研究成果，获得（　　　　　　　　）课题（组）经费或实验室的资助，在（　　　　　　　）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

年　　月　　日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（　　）1.经厦门大学保密委员会审查核定的保密学位论文，于　年　月　日解密，解密后适用上述授权。

（　　）2.不保密，适用上述授权。

（请在以上相应括号内打"√"或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年　月　日

摘要

状态空间模型和马尔可夫体制转换模型是计量经济学中两种富有成效的模型，本文着重介绍测量矩阵是隐马尔可夫链的状态空间模型。在经典统计方法下，对未观测状态向量或马尔可夫转换变量的推断是基于模型的参数估计，这种方法忽视了参数估计的不确定性或者抽样的变化。本文分别从经典统计和贝叶斯统计出发考虑参数估计的不确定性问题。主要工作在于下面三点：

1. 在经典统计下，改进了Stoffer和Wall(1991)在状态空间模型下的Bootstrap方法，来估计参数的抽样分布，使之适用于本文所述模型。

2. 在贝叶斯方法下，假设参数是随机向量，自然而然的结合了参数的不确定性。首先介绍Gibbs抽样来获得参数后验估计。最后简单回顾SMC方法，给出了适用于本文所述模型的SMC参数学习算法。

3. 对本文所述三种方法做了模拟实验，并应用于美国月度肺炎和流感死亡率数据。

关键词：状态空间模型；马尔可夫体制转换；自助抽样；Gibbs抽样；序贯蒙特卡洛。

Abstract

State-space models and Markov-switching models have both been highly productive paths for research in econometrics, in this text, we focus on the state-space models with markov-switching in measurement matrix. In the classical approach, inference on the unobserved state vector or the Markov-switching variable is made conditional on the parameter estimates of the model. This approach ignores uncertainty or the sampling variation associated with the parameter estimates of the model. In this paper, we take classical approach and bayes approach to consider parameter uncertainty problem. The main work are summarized as follows:

1. In classical approach, develop Stoffer and Wall method, which used in state space model, to our model.

2. In bayes approach, the parameters are assumed to be random vectors, so it's posterior distribution incorporates uncertainty naturally. First, Gibbs sample are introduced to get parameter posterior distribution. Finally, we review Sequential Monte Carlo method, and propose a parameter learning algorithm to our model.

3. Three methods are tested against a synthetic time series, and applied to U.S. monthly pneumonia and influenza mortality data.

Key Words: State-space model, Markov-switching model, Bootstrap, Gibbs sample, sequential Montr Carlo.

目 录

摘要 I

Abstract II

第一章 引言 1

1.1 研究背景 1

1.2 本文的主要工作 2

1.3 本文的结构 3

第二章 状态空间模型 4

2.1 状态空间模型及其估计 4

2.2 状态空间模型bootstrap程序 6

2.3 序贯蒙特卡洛 8

2.3.1 粒子滤波基础知识 8

2.3.2 参数学习 10

第三章 带马尔可夫转换的状态空间模型 12

3.1 模型的滤波及估计 12

3.2 Bootstrap模型参数分布 15

3.3 Gibbs 抽样 17

3.4 SMC参数学习 18

第四章 模拟实验及应用 20
4.1 模拟实验 20
4.2 美国月度肺炎和流感死亡率数据分析 26
第五章 总结和展望 31
参考文献 32
致谢 35

第一章　引言

1.1 研究背景

状态空间模型(State-space models)和马尔可夫转换模型(Markov-switching models)是计量经济学研究中两种富有成效的模型，它们都刻画了在我们尝试去理解经济行为时面临的一些基本问题。未观测变量在消费者行为，失业率，通货膨胀动态变化，经济活动指数，货币政策和金融市场中都扮有重要的角色。在这些情形下，通过卡尔曼滤波(Kalman filter)运算的状态空间框架，是我们在时间序列分析中用来做统计推断的唯一方法。同时也有引人注目的经验证据，表明经济系统表现为从一个体制到另一个体制的偶然跳跃。当这样一个转变发生时，数据的分布似乎也要改变。例如，宏观经济周期性地从繁荣到衰退，再从衰退回到繁荣，这两个体制之间的变化是不同的。金融市场周期性地从低波动率体制到高波动率体制，再从高波动率回到低波动率。用马尔可夫过程(Markov process)来刻画这些变化是吸引人的。

正是基于上面的描述，本文考虑线性高斯转换状态空间模型(Linear Gaussian Switching State-Space Models, SSSM)。它普遍存在于统计学(Cappé等, 2005；Frühwirth-Schnatter, 2006)，经济学(Kim和Nelson, 1999；Giordani等, 2007)和信号处理(Barembruch等, 2009；Costa等, 2005)中。SSSM已经成功的被用于描述多变点模型(Fearnhead和Clifford, 2003；Giordani和Kohn, 2008)，有溢出点的非参数回归模型(Carter和Kohn, 1996)和马尔可夫自回归模型(Billio和Monfort, 1998；Frühwirth-Schnatter, 2006；Kim和Nelson, 1999)。

在状态空间模型下，通过卡尔曼滤波计算一步向前预测误差，然后建立高斯似然函数，最后用极大似然法来获得参数估计。在通常的应用中，使用MLE递归地计算滤波与平滑，面临着恰当地考虑参数不确定性的困难。由于存在的渐进理论，在状态空间模型下通过卡尔曼滤波估计的推断问题是易于处理的。在适当的条件下，通过最大似然方法获得的参数估计和通过卡尔曼滤波得到状态估计已经被证实是相合和渐进正态的(Ljung和Caines,1979；Spal和Wall,1984)。然而时间序列数据长度经常比适度的长度短，这样渐进理论的使用是值得可疑的。对于ARMA模型，一些学者已经发现证据，样本必须相当大，渐进结论才

是可用的(Dent和Min,1978；Ansley和Newbold,1980)。还有，当参数接近参数空间的边界时，问题也会发生，这样类似的情况也发生在状态空间模型中。Stoffer和Wall(1991)对这一问题，采用了Efron(1979)提出的非参数蒙特卡罗(Monte Carlo)Bootstrap方法，证实了上述的问题，以及在小样本下Bootstrap方法的优越性和大样本下与渐进理论的一致性。

统计学中另一个领域，贝叶斯方法(Bayesian approach)对参数不确定性问题提供了更加一致的看法，因为未知参数被看作是一个随机向量。贝叶斯方法将先验信息和样本数据相结合，得到参数的后验分布，然后在后验分布的基础上进行统计推断。然而后验分布往往因为积分过于复杂，得不到后验分布的完全形式，更多时候得到的是后验分布的核。得益于计算机和计算技术的发展，模拟来自于一个马尔可夫链的随机数的蒙特卡洛方法，叫做马尔可夫链蒙特卡洛(Markov Chain Monte Carlo,MCMC)方法，是现在进行贝叶斯数据分析的标准方法，并且得到越来越广泛的应用。

然而，MCMC方法不能用于在线推断，因为当一个新的观测值实现时，需要去模拟产生一个完全新的马尔可夫链。换句话说，在MCMC方法下，基于t-1个观测值的输出不能用来计算t个观测值时的后验分布。在这个意义下，MCMC不像卡尔曼滤波，不能顺序地使用。最近十几年快速发展的一种模拟方法，叫做序贯蒙特卡洛(Sequential Monte Carlo,SMC)，已经被证明成功的用于在线滤波，在有未知参数的动态线性模型和一般的非线性非高斯状态空间模型。SMC提供了另外一种基于模拟的方法来近似复杂的后验分布。

1.2 本文的主要工作

本文首先介绍状态空间模型，接着引出一类SSSM模型，具体的是观测矩阵是隐马尔科夫链的状态空间模型，从经典统计和贝叶斯统计两个角度来考虑参数不确定性问题。

与现有的文献相比，本文的主要工作在于下面三点：

第一、由于经典统计下MLE的渐进正态性面临着样本量和参数空间边界的问题，改进了Stoffer和Wall(1991)在状态空间模型下的Bootstrap方法来估计参数的抽样分布，使之适用于本文所述模型。

第二、在贝叶斯方法下，结合第三章模型所述的滤波概率下，利用Gibbs抽样来获得参数后验估计。简单回顾SMC方法，给出了适用于本文所述模型

的SMC参数学习算法。

第三、对本文所述三种方法做了模拟实验，并应用于美国月度肺炎和流感死亡率数据。

1.3 本文的结构

本文的结构安排如下:

第一章,介绍研究背景，简单说明本文所做的主要工作，给出本文的结构和轮廓。

第二章,介绍状态空间模型，卡尔曼滤波与平滑，以及相应的渐近理论和bootstrap程序，最后介绍了SMC方法，为下一章内容做基础与铺垫。

第三章,介绍带体制转换的状态空间模型以及滤波，之后引入相应的Bootstrap抽样程序，Gibbs抽样程序以及SMC方法来估计参数。

第四章,进行模拟实验，证明方法的有效性，最后应用到实际数据进行实证分析。

第五章,总结和展望。

第二章　状态空间模型

本章首先介绍状态空间模型及其一些性质，接着介绍Stoffer和Wall(1991)提出的Bootstrap抽样，最后介绍SMC方法。在下一章，我们把它们推广到带马尔可夫体制转换的状态空间模型。

2.1 状态空间模型及其估计

状态空间模型或动态线性模型通过如下的方程定义

$$\boldsymbol{x}_{t+1} = \Phi\boldsymbol{x}_t + \Upsilon\boldsymbol{u}_t + \boldsymbol{w}_t \quad t = 0, 1, ..., n \tag{2.1}$$

$$\boldsymbol{y}_t = A_t\boldsymbol{x}_t + \Gamma\boldsymbol{u}_t + \boldsymbol{v}_t \quad t = 1, ..., n. \tag{2.2}$$

这里$\boldsymbol{x}_t$代表了p维的未观测的状态向量，$\boldsymbol{y}_t$是一个q维的观测向量。在状态方程2.1中，初始状态$\boldsymbol{x}_0$是均值为$\boldsymbol{\mu}_0$和协方差矩阵为Σ_0的正态向量，Φ是$p \times p$维矩阵，Υ是$p \times r$维矩阵，$\boldsymbol{u}_t$是$r{\times}1$维的确定的输入向量。在观测方程2.2中，A_t是$q \times p$维的观测矩阵，Γ是$q \times r$维矩阵。这里$\boldsymbol{w}_t$和$\boldsymbol{v}_t$是白噪声序列，且都和$\boldsymbol{x}_0$独立，$\mathrm{var}(\boldsymbol{w}_t) = Q$，$\mathrm{var}(\boldsymbol{v}_t) = R$，我们还允许状态噪音和观测噪音在时刻$t$是相关的：也就是$\mathrm{cov}(\boldsymbol{w}_t, \boldsymbol{v}_t) = S$,而在其它时刻等于零。我们假设模型的系数和模型的相关结构被一个$k \times 1$维的参数向量θ 唯一的参数化；也就是$\Phi = \Phi(\theta)$，$\Upsilon = \Upsilon(\theta)$，$Q = Q(\theta)$，$\Gamma = \Gamma(\theta)$，$R = R(\theta)$，$S = S(\theta)$。

从实际应用的角度，状态空间模型分析的主要目的是获得潜在未知信号$\boldsymbol{x}_t$的估计值，在给定直到时刻s的数据$\boldsymbol{Y}_s = \{\boldsymbol{y}_1, ..., \boldsymbol{y}_s\}$。当$s < t$，这个问题叫做预测。当$s = t$时，这个问题叫滤波，以及$s > t$时，叫平滑。获得这些估计值的同时，我们还需要知道估计的精度。

在本文中，我们用到的记号定义如下：

$$\boldsymbol{x}_t^s = E(\boldsymbol{x}_t|Y_s) \tag{2.3}$$

$$P_{t_1,t_2}^s = E\{(\boldsymbol{x}_{t_1} - \boldsymbol{x}_{t_1}^s)(\boldsymbol{x}_{t_2} - \boldsymbol{x}_{t_2}^s)'\}. \tag{2.4}$$

在2.4中当$t_1 = t_2 = t$时,为了方便记为P_t^s。

从而我们定义$\boldsymbol{x}_{t+1}$在给定数据$\{\boldsymbol{y}_1, ..., \boldsymbol{y}_t\}$下的最佳线性预测为$\boldsymbol{x}_{t+1}^t$，以及预测误差$(\boldsymbol{x}_{t+1} - \boldsymbol{x}_{t+1}^t)$的协方差矩阵为$P_{t+1}^t$。卡尔曼滤波可以用来相继地获得预测值和它们的协方差矩阵，当新的观测值实现时。新息序列，$\{\epsilon_t, t = 1, ..., n\}$ 定

义为在给定数据$\{\boldsymbol{y}_1, ..., \boldsymbol{y}_{t-1}\}$下$\boldsymbol{y}_t$的最佳线性预测的误差序列。新息为

$$\boldsymbol{\epsilon}_t = \boldsymbol{y}_t - E(\boldsymbol{y}_t|Y_{t-1}) = \boldsymbol{y}_t - \boldsymbol{A}_t\boldsymbol{x}_t^{t-1} - \Gamma\boldsymbol{u}_t, \tag{2.5}$$

对$t = 1, ..., n$。注意，$E(\boldsymbol{\epsilon}_t) = \mathbf{0}$以及新息协方差矩阵为

$$\Sigma_t = var(\boldsymbol{\epsilon}_t) = A_tP_t^{t-1}A_t' + R. \tag{2.6}$$

性质 2.1: 卡尔曼滤波

状态空间模型2.1和2.2，在初始条件$\boldsymbol{x}_1^0 = \Phi\boldsymbol{\mu}_0 + \Upsilon\boldsymbol{u}_0$和$P_1^0 = \Phi\Sigma_0\Phi' + Q$下，对$t = 1, ..., n$的卡尔曼滤波的新息形式，由下面的等式给出：

$$\begin{aligned}
\boldsymbol{x}_{t+1}^t &= \Phi\boldsymbol{x}_t^{t-1} + \Upsilon\boldsymbol{u}_t + K_t\boldsymbol{\epsilon}_t, && (2.7)\\
P_{t+1}^t &= \Phi P_t^{t-1}\Phi' + Q - K_t\Sigma_tK_t', && (2.8)\\
K_t &= (\Phi P_t^{t-1}A_t' + S)\Sigma_t^{-1}, && (2.9)\\
\boldsymbol{x}_t^t &= \boldsymbol{x}_t^{t-1} + P_t^{t-1}A_t'\Sigma_t^{-1}\boldsymbol{\epsilon}_t, && (2.10)\\
P_t^t &= P_t^{t-1} - P_t^{t-1}A_t'\Sigma_t^{-1}A_tP_t^{t-1} && (2.11)
\end{aligned}$$

性质 2.2: 卡尔曼平滑

对状态空间模型2.1和2.2，初始条件x_n^n和P_n^n由性质2.1获得，在$t = n, n-1, ..., 1,$下

$$\begin{aligned}
x_{t-1}^n &= \boldsymbol{x}_{t-1}^{t-1} + J_{t-1}(\boldsymbol{x}_t^n - \boldsymbol{x}_t^{t-1}), && (2.12)\\
P_{t-1}^n &= P_{t-1}^{t-1} + J_{t-1}(P_t^n - P_t^{t-1})J_{t-1}', && (2.13)\\
J_{t-1} &= P_{t-1}^{t-1}\Phi'[P_t^{t-1}]^{-1}. && (2.14)
\end{aligned}$$

模型参数θ的估计可以通过高斯拟最大似然函数求得。忽略常数部分，新息形式的高斯似然函数是

$$-\ln L_Y(\theta) = \frac{1}{2}\sum_{t=1}^{n}(\ln|\Sigma_t(\theta)| + \boldsymbol{\epsilon}_t(\theta)'\Sigma_t(\theta)^{-1}\epsilon_t(\theta)) \tag{2.15}$$

这里$L_Y(\theta)$定义为在正态假设下，给定$\boldsymbol{y}_1, ..., \boldsymbol{y}_n$的$\theta$的似然函数。最大化2.15关于$\theta$产生高斯最大似然估计$\hat{\theta}$=argmax$_\theta L_Y(\theta)$。很明显2.15是一个关于未知参数的非线性和复杂的函数，存在很多形式的迭代程序来最大化这个函数以及较容易实施(Burmeister和Wall,1982)。

运用Newton-Raphson估计程序的步骤如下：

1.选择一个初始参数，记为$\theta^{(0)}$。

2.使用初始参数值$\theta^{(0)}$去获取卡尔曼滤波以及新息和误差方差，即$\{\boldsymbol{\epsilon}_t^{(0)}; t = 1, ..., n\}$ 和$\{\Sigma_t^{(0)}; t = 1, ..., n\}$。

3.以$-\ln L_Y(\theta)$作为准则函数，用Newtow-Raphson程序进行一步迭代，获得一个新的估计，记为$\theta^{(1)}$.

4.在第j步迭代($j = 1, 2, ...$)，重复步骤2 用$\theta^{(j)}$替代$\theta^{(j-1)}$，得到新的新息值$\{\boldsymbol{\epsilon}_t^{(j)}; t = 1, ..., n\}$和$\{\Sigma_t^{(j)}; t = 1, ..., n\}$。然后重复步骤3去获得一个新的估计$\theta^{(j+1)}$，直到相邻两次估计值或者似然函数之差小于一个预先给定的较小的值时，停止迭代。

在Caines(1988,Chapters 7 and 8)以及Hannan和Deistler(1988,Chapter 4)中讨论了模型参数估计$\widehat{\theta}_n$ 的渐进分布，在这两个参考文献里，都在一般条件下建立了估计的相合和渐进正态性。我们假设真实的参数是θ_0，以及θ_0的维数就是参数空间的维数。我们有如下的性质。

性质 2.3: 估计的渐进分布

在一般条件下，记$\hat{\theta}_n$是θ_0最大似然估计。当$n \longrightarrow \infty$，

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1}) \tag{2.16}$$

这儿$I(\theta)$是渐进信息矩阵，

$$I(\theta) = \lim_{n\to\infty} \frac{1}{n} E[-\partial^2 \ln L_Y(\theta)/\partial(\theta)\partial\theta']. \tag{2.17}$$

精确的细节和证明在上述的两篇参考书里。对Newton程序，在收敛时刻的Hessian矩阵可以用来作为$nI(\theta_0)$的估计。即：

$$\widehat{I_n(\theta_0)} = -\frac{1}{n}\frac{\partial^2 \ln L_Y(\theta)}{\partial(\theta)\partial\theta'}|_{\theta=\hat{\theta}} \tag{2.18}$$

相应的$\hat{\theta}$估计的协方差矩阵为

$$\widehat{Var}(\hat{\theta}_n) = \frac{1}{n}[\widehat{I_n(\theta_0)}]^{-1} \tag{2.19}$$

2.2 状态空间模型bootstrap程序

除了依赖于渐进标准误，参数估计的分布可以通过bootstrap获得。这个方

法可能比用渐进结论更优，特别是在中小样本情形下，或者新息的分布不是正态的。这个程序由Stoffer和Wall(1991)提出，概括如下。

我们假设模型的参数估计已经获得，以及在$\theta=\hat{\theta}$下的卡尔曼滤波也已经得到，这样估计的新息$\epsilon_t(\theta)$也有了。一旦这些已经得到，bootstrap程序由如下的5步构成。

1.构造标准化新息，通过

$$\boldsymbol{e}_t(\hat{\theta})=\Sigma_t^{-1/2}(\hat{\theta})\boldsymbol{\epsilon}_t(\hat{\theta}) \tag{2.20}$$

这里$\Sigma_t^{-1/2}(\theta)$是$\Sigma_t(\theta)$矩阵的唯一的逆平方根矩阵。通过上式我们使得模型的所有残差最起码有相同的一阶，二阶矩。

2.从集合$\{\boldsymbol{e}_1(\hat{\theta}),...,\boldsymbol{e}_n(\hat{\theta})\}$中有放回的抽样$n$次，得到一个标准化了的新息的bootstrap样本$\{\boldsymbol{e}_1^*,...,\boldsymbol{e}_n^*\}$。

3.使用新息表达式，构造bootstrap数据集$\{\boldsymbol{y}^*(t);1\leqslant t\leqslant n\}$如下：定义$(p+q)\times 1$向量

$$\begin{bmatrix}\boldsymbol{x}_{t+1}^t\\ \boldsymbol{y}_t\end{bmatrix}$$

结合2.5和2.7得到向量$\boldsymbol{\xi}_t$的一阶方程

$$\boldsymbol{\xi}_t=F_t\boldsymbol{\xi}_{t-1}+G\boldsymbol{u}_t+H_t\boldsymbol{e}_t, \tag{2.21}$$

这里

$$F_t=\begin{bmatrix}\Phi & 0\\ A_t & 0\end{bmatrix},G=\begin{bmatrix}\Upsilon\\ \Gamma\end{bmatrix},H_t=\begin{bmatrix}K_t\Sigma_t^{1/2}\\ \Sigma_t^{1/2}\end{bmatrix}. \tag{2.22}$$

这样解2.21用$\boldsymbol{e}_t^*$取代$\boldsymbol{e}_t$，就可以得到$\{\boldsymbol{y}^*(t);1\leqslant t\leqslant n\}$。外生变量$\boldsymbol{u}_t$以及卡尔曼滤波的初始条件仍旧固定在它们的给定值，参数向量固定在$\hat{\theta}$。因为初始的不规则，有时让$\boldsymbol{y}_t^*\equiv\boldsymbol{y}_t$，$t=1,2,...t_0$，是一个好主意，这里$t_0$是一个比较小的值，然后从$\{\boldsymbol{e}_{t_0+1}(\hat{\theta}),...,\boldsymbol{e}_n(\hat{\theta})\}$中抽样。也就是不要bootstrap开始得几个数据点，一般让t_0等于4或5就足够了。

4.用bootstrap数据集$\{\boldsymbol{y}_t^*;t=1,...,n\}$,构造似然函数$L_{Y^*}(\theta)$, 获得$\theta$的MLE估计，记为$\hat{\theta}^*$

5.重复步骤2-4 B次，B是一个比较大的数，从而获得关于参数估计的一个bootstrap集$\{\hat{\theta}_b^*;b=1,...,B\}$。最后用$\hat{\theta}_b^*$的分布去估计$\hat{\theta}$的分布。

2.3 序贯蒙特卡洛

2.3.1 粒子滤波基础知识

粒子滤波(Particle filter)，是SMC在状态空间模型中的一种叫法。让我们考虑更加一般的动态模型，这里放宽了正态假设以及线性假设。观测方程和状态转移方程变为

$$y_t|x_t \sim p(y_t|x_t) \tag{2.23}$$
$$x_t|x_{t-1} \sim p(x_t|x_{t-1}), t = 1, 2, ... \tag{2.24}$$

定义初始状态概率密度$p(x_0)$。所有的参数在本小节里假设为已知。式2.7和2.10的卡尔曼递归现在分别为

$$p(x_t|Y_{t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|Y_{t-1})dx_{t-1} \tag{2.25}$$
$$p(x_t|Y_t) = \frac{p(y_t|x_t)p(x_t|Y_{t-1})}{p(y_t|Y_{t-1})} \tag{2.26}$$

除了在正态动态线性模型，式2.25关于x_{t-1}的积分和式2.26贝叶斯公式都是不能解析的获得或者计算昂贵的。宽松地讲，粒子滤波结合了卡尔曼滤波相继地估计的特性以及MCMC抽样对模型的灵活性，同时避免了它们的一些缺点。一方面，像MCMC抽样而不是卡尔曼滤波，粒子滤波允许更加灵活的观测和动态演化方程以及分布。另一方面，像卡尔曼滤波而不是MCMC抽样，粒子滤波提供了对状态和参数的实时滤波和平滑。

大多数粒子滤波的目标是，在给定初始独立同分布的粒子集$\{x_{t-1}^{(i)}\}_{i=1}^N$近似$p(x_{t-1}|Y_{t-1})$下，获得独立同分布的粒子集$\{x_t^{(i)}\}_{i=1}^N$近似$p(x_t|Y_t)$。

最常见的滤波是Gordon等(1993)提出的自助滤波(bootstrap filter,BF)，以及Pitt和Shephard(1999)提出的辅助粒子滤波(auxiliary particle filter,APF)。Lopes和Tsay (2011) 认为BF和APF界定了粒子滤波文献中两种根本性不同的方向，即抽样-重抽样(sample-resample)和重抽样-抽样(resample-sample)方法。叙述如下：

传递-重抽样滤波(Propagate-resample filters)

Gordon(1993)等提出的BF是基于序贯抽样重要重抽样(sample importance resample,SIR)的思想。卡尔曼递归式2.25和2.26组合为

$$p(x_t, x_{t-1}|y_t, Y_{t-1}) \propto \underbrace{p(y_t|x_t)}_{2.Resample} \underbrace{P(x_t|x_{t-1})P(x_{t-1}|Y_{t-1})}_{1.Propagate} \tag{2.27}$$

换句话说，BF首先传递(propagate)时刻$t-1$的后验分布的粒子，来得到来自时刻t的先验分布的粒子。然后将传递得到的粒子重抽样，权重和它们的似然函数成正比。这就是下面的算法1。

重抽样-传递滤波(Resample-propagate filters)

相似地，APF首先对时刻$t-1$后验分布的粒子重抽样，考虑到下一个观测值y_t并计入权重。然后，把重新抽样的粒子传递到时刻t。同式2.27一样，重写为

$$p(x_t, x_{t-1}|y_t, Y_{t-1}) \propto \underbrace{P(x_t|x_{t-1}, Y_t)}_{2.Propagate} \underbrace{p(y_t|x_{t-1})P(x_{t-1}|Y_{t-1})}_{1.Resample} \tag{2.28}$$

APF的主要困难是，在大多应用中，$p(y_t|x_{t-1})$逐点计算不可得或者从$p(x_t|x_{t-1}, Y_t)$的抽样不容易实现。Pitt和Shephard(1999)建议：

(a) 用$p(y_t|g(x_{t-1}))$，也就是$p(y_t|x_t)$在$g(x_{t-1})$(通常使用状态转移密度$p(x_t|x_{t-1})$的期望值，中位数，众数)的密度，作为建议权重(proposal weight)对旧的粒子x_{t-1}重新抽样。

(b) 用$q(x_t|x_{t-1}, y_t) \equiv p(x_t|x_{t-1})$作为建议密度(proposal density)去传递重抽样产生的粒子到新的粒子集$\{x_t^{(i)}\}_{i=1}^N$。注意这里$q(\cdot)$是盲目的，因为没有考虑到现在的观测值y_t。

因为重抽样和传递粒子都是来自建议密度，所以产生的粒子有相应的权重

$$w_t \propto \frac{p(y_t|x_t)p(x_t|x_{t-1})p(x_{t-1}|Y_{t-1})}{p(y_t|g(x_{t-1}))p(x_t|x_{t-1})p(x_{t-1}|Y_{t-1})} \tag{2.29}$$

$$= \frac{p(y_t|x_t)}{p(y_t|g(x_{t-1}))} \tag{2.30}$$

这就是下面的算法2。

算法1：Bootstrap filter (BF)

1. 传递$\{x_{t-1}^{(i)}\}_{i=1}^N$到$\{\tilde{x}_t^{(i)}\}_{i=1}^N$通过$p(x_t|x_{t-1})$
2. 从$\{\tilde{x}_t^{(i)}\}_{i=1}^N$以权重$w_t^{(i)} \propto p(y_t|\tilde{x}_t^{(i)})$重抽样产生$\{x_t^{(i)}\}_{i=1}^N$。

算法2：Auxiliary particle filter (APF)

1. 从$\{x_{t-1}^{(i)}\}_{i=1}^N$以权重$w_t^{(i)} \propto p(y_t|g(x_{t-1}^{(i)}))$重抽样产生$\{\tilde{x}_{t-1}^{(i)}\}$